Genome-wide association for metabolic clusters in early-lactation Holstein dairy cows

Genome-wide association for metabolic clusters in early-lactation Holstein dairy cows

Journal:	*Journal of Dairy Science*
Manuscript ID	JDS.2019-17369.R2
Article Type:	Research
Date Submitted by the Author:	12-Jan-2020
Complete List of Authors:	Atashi, Hadi; Ghent University, Department of Reproduction, Obstetrics and Herd Health; Shiraz University, Department of Animal Science Salavati, Mazdak; The University of Edinburgh, The Roslin Institute and Royal (Dick) School of Veterinary Studies De Koster, Jenne; Ghent University, Department of Reproduction, Obstetrics and Herd Health Crowe, Mark; University College Dublin, Animal Husbandry & Production Opsomer, Geert; Ghent University, Department of Reproduction, Obstetrics and Herd Health Hostens, Miel; Ghent University, Department of Reproduction, Obstetrics, Herd Health
Key Words:	Lactation, Milk Synthesis

Running Title: Genome-wide association for metabolic clusters in early-lactation dairy cows

Title: Genome-wide association for metabolic clusters in early-lactation Holstein dairy cows

H. Atashi*§, M. Salavati†, J. De Koster*, M. A. Crowe‡, G. Opsomer*, the GplusE consortium and M. Hostens*[1]

*Department of Reproduction, Obstetrics and Herd Health, Ghent University, Merelbeke, Belgium.

†The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Midlothian, UK.

‡University College Dublin, Dublin, Ireland.

§Department of Animal Science, Shiraz University, Shiraz, Iran.

[1] Corresponding author: Miel Hostens, Salisburylaan 133, 9820 Merelbeke, Belgium.

Phone number: +32478593703

E-mail: miel.hostens@ugent.be

INTERPRETIVE SUMMARY

The aim of this study was to detect genomic region(s) associated with metabolic clusters in early-lactation Holstein cows. Blood metabolites (experiment I) and Fourier-transform mid-infrared (FT-MIR) spectra of milk (experiment II) were used to predict metabolic clusters (BALANCED vs. OTHER) for, respectively, 105 and 4,267 Holstein dairy cows. The results of experiment I revealed eight SNPs (on BTA11, BTA23 and BTAX) associated with the predicted metabolic clusters. Considering the results of experiment II, it can be concluded that metabolic clusters predicted based on FT-MIR spectra in milk samples collected during early lactation, is a highly polygenic trait regulated by many small-sized effects. The heritability for the predicted metabolic clusters was 0.17, indicating that the use of FT-MIR spectra in milk to predict metabolic clusters in early-lactation across a large number of cows has a satisfactory potential to be included in the genetic selection program of modern dairy cows.

GENOME-WIDE ASSOCIATION FOR METABOLIC CLUSTERS IN EARLY-LACTATION HOLSTEIN DAIRY COWS

ABSTRACT

The aim of this study was to detect genomic region(s) associated with metabolic clusters in early-lactation Holstein cows. This study was carried out in two experiments. In experiment I, which was carried out on 105 multiparous Holstein cows, k-means clustering on log-transformed and standardized concentrations of blood glucose, insulin like growth factor I, free fatty acids, and β-hydroxybutyrate at 14 and 35 days in milk (**DIM**) was used to classify animals in metabolic clusters (BALANCED vs. OTHER). Forty percent of the animals were categorized in the BALANCED group, while the remainder ended up in the OTHER category. The cows were genotyped for a total of 777,962 single nucleotide polymorphisms (**SNPs**). A genome-wide association study was done using a case–control approach through the GEMMA software accounting for population structure. There were eight SNPs (BTA11, BTA23 and BTAX) associated with the predicted metabolic clusters. In experiment II, carried out on 4,267 second-parity Holstein cows, milk samples collected starting from the first week until 50 DIM were used to determine Fourier-transform mid-infrared (**FT-MIR**) spectra and subsequently to classify the animals in the same metabolic clusters (BALANCED vs. OTHER). Twenty-eight percent of the animals were categorized in the BALANCED group, while the remainder ended up in the OTHER category. Although daily milk yield was lower in BALANCED cows, there was no difference in daily fat- and protein-corrected milk yield in cows from the BALANCED metabolic cluster compared with those belonging to the OTHER metabolic cluster. In the next step, a single-step genomic BLUP was used to identify genomic region(s) associated with the predicted metabolic clusters. The

results revealed that the prediction of metabolic clusters is a highly polygenic trait regulated by many small-sized effects. The region of 36,258 to 36,295 kb on BTA27 was the highly associated region for the predicted metabolic clusters, with the closest genes to this region (*ANK1* and *miR-486*) being related to hematopoiesis, erythropoiesis, and mammary gland development. The heritability (SD) for metabolic clustering was 0.17 (0.03), indicating that the use of FT-MIR spectra in milk to predict metabolic clusters in early-lactation across a large number of cows has a satisfactory potential to be included in the genetic selection program of modern dairy cows.

Key words: metabolic adaptation, transition period, genome-wide association study, dairy cow

INTRODUCTION

The transition period, defined as the last three weeks before and the first three weeks after calving, is critically important to health, production, and profitability of dairy cows (Drackley, 1999, LeBlanc, 2010). The early-lactation cow needs comprehensive adaptive reactions to cope with the challenges caused by the transition from the pregnant, non-lactating state to the non-pregnant, lactating state (Bell, 1995, Goff and Horst, 1997, Esposito *et al*., 2014). Dry matter intake (**DMI**) starts to decrease a few weeks before parturition where its lowest level occurs at calving (Bell, 1995, Ingvartsen and Andersen, 2000). Although most cows rapidly increase DMI in the first weeks after calving (Ingvartsen and Andersen, 2000), high-yielding dairy cows enter a state of negative energy balance (**NEB**) around calving when the energy demand for maintenance and lactation exceeds the dietary intake potential (Bauman and Currie, 1980, Patton *et al*., 2006, Alawneh *et al*., 2012). The NEB results in a mobilization of body reserves to provide energy requirements for milk synthesis (Van der Drift *et al*., 2012). Although lipid mobilization is a normal and necessary process in the metabolic adaptation to support lactation, many health disorders are associated with uncontrolled lipid mobilization (Contreras *et al*., 2018). Increased levels of metabolic and infectious disorders (mastitis, ketosis, fatty liver, hypocalcemia, and retained placenta) and decreased reproductive performance (anestrus, decreased conception rate, and delayed reproduction) have been linked with longer and more severe periods of NEB (Drackley, 1999, Collard *et al*., 2000, LeBlanc, 2010).

The ability of the animal to cope with the NEB in the transition period varies considerably between individuals and may be partly explained by the genetic background of animals

(Kessel *et al*., 2008, Liinamo *et al*., 2012, Spurlock *et al*., 2012, Pryce *et al*., 2016). Although, some countries have already initiated genetic evaluation of metabolic disorders, there are opportunities to use predictor traits and genomic information to strengthen genetic evaluations for metabolic health in dairy cows. Challenges such as lack of a standard recording system and an accurate data collection, difficulty in diagnosing subclinical cases and in determining economic values for metabolic disorders, are among the reasons why interest is growing in using easily measurable predictors of metabolic diseases, either recorded on-farm by using sensors and milk tests or off-farm using data collected from routine milk recording (Zwald *et al*., 2004, Pryce *et al*., 2016). Clustering early-lactation cows based on different blood or milk metabolites can be used as an effective way to group individual cows based on their ability to cope with the altered metabolic challenges of lactation (De Koster *et al*., 2019). This clustering methodology can be used as a diagnostic or a herd management tool but, on a larger scale, can also be used to identify superior animals with a high genetic value for targeted breeding purposes. Metabolic imbalance is defined as 'a condition where the regulatory mechanisms are insufficient for the animals to function optimally leading to a high risk of digestive, metabolic and infectious problems' (Ingvartsen, 2006). Elevated levels of free fatty acids (**FFA**) and β-hydroxybutyrate (**BHB**), and decreased levels of glucose and insulin like growth factor I (**IGF-I**) are reported as blood biomarkers of metabolically imbalanced cows (Ingvartsen et al., 2003, Puppel and Kuczynska, 2016).

Although physiological adaptations of dairy cows in the transition period are well described, knowledge about the genetic background still remains insufficient (Bell and Bauman, 1997,

Drackley *et al*., 2005). Genome-wide association studies (**GWAS**) provide a fine mapping strategy for underlying genetic background of complex traits in humans, animals and plants (Zhang *et al*., 2012, Wang *et al*., 2015a). The aim of the present study was to use GWAS to identify genetic markers linked with metabolic clusters in early-lactation Holstein cows. In order to predict metabolic clusters of animals we used blood metabolites (experiment I) and Fourier-transform mid-infrared (**FT-MIR**) spectra of milk (experiment II).

MATERIALS AND METHODS

EXPERIMENT I

Animals and phenotype

The data in experiment I were collected as part of the Genotype plus Environment (**GplusE**) FP7-Project (http://www.gpluse.eu). Detailed description of the experiments, laboratory analysis, phenotypic data and the k-means clustering methodology are explained in De Koster *et al*. (2019). In short, a total of 105 multiparous Holstein dairy cows from four research herds (Aarhus University, Denmark; UCD Lyons Research farm, University College Dublin, Ireland; Agri-Food and Biosciences Institute, Northern Ireland, UK; and Leibniz Institute for Farm Animal Biology, Germany) were combined. Glucose, IGF-1, FFA and BHB concentrations were determined in blood samples taken around 14 and 35 DIM as described in Foldager *et al*. (2019). K-means clustering based on the log-transformed and standardized blood metabolite concentrations (mean = 0 and SD = 1) was used to describe metabolic profile of the included cows. The cluster of cows with high concentrations of glucose and IGF-I and low concentrations of FFA and BHB at 14 and 35 DIM was defined

as a BALANCED metabolic cluster and was compared with OTHER. The mean glucose (m*M*), IGF-I (ng/mL), FFA (m*M*) and BHB (m*M*) values for BALANCED cows at 14 DIM were 3.62, 121.66, 0.52, and 0.44, respectively. The corresponding values for cows belonging to the OTHER cluster were 3.28, 49.63, 0.81, and 0.71, respectively.

Genome-wide association study

All cows were genotyped using Illumina BovineHD BeadChip for a total of 777,962 SNPs (San Diego, CA, USA). Quality control for markers was done using PLINK (Purcell *et al*., 2007). SNPs with no position information, call rate less than 95%, minor allele frequency less than 5% or a Hardy-Weinberg equilibrium *P* value less than 6.4×10^{-8} were excluded. Single SNP association analysis was performed using a mixed model approach in GEMMA software (Zhou and Stephens, 2012). GEMMA implements the genome-wide efficient mixed model association algorithm and accounts for population stratification. The centered relatedness matrix was calculated from all genotypes. The association test was then performed with phenotype, genotype and the centered relatedness matrix files using a univariate linear mixed model. Each SNP was fitted as a covariate and the likelihood ratio test was used for each SNP against the null hypothesis of g = 0 using the following statistical model.

$$y = W\alpha + x\beta + u + \varepsilon$$

Where **y** is a n × 1 vector of phenotype values for all individuals (the predicted metabolic clusters and concentrations of BHB, glucose, IGF-1, and FFA); **W** is an n × c matrix of covariates (fixed effects containing herd, parity and the mean); **α** is an c × 1 vector of the

corresponding coefficients including the intercept; **x** is an n × 1 vector of genotypes of a marker at the locus tested; β is the effect size of the marker; **u** is a n × 1 vector of random polygenic effects with a covariance structure as **u**~ N(0, **K**Vg), where **K** is an n × n marker based additive genetic relatedness matrix and Vg is the polygenic additive variance; and **ε** is an n × 1 vector of residual errors with **ε**~ N(0, **I**Ve), where **I** is an n × n identity matrix and Ve is the residual variance. The thresholds of the Bonferrooni corrected *P* values for suggestive and 5% genome-wide significance association were set as 1.74×10^{-6} (1 divided by the number of SNPs) and 8.68×10^{-6} (0.05 divided by the number of SNPs), respectively (Lander and Kruglyak, 1995).

EXPERIMENT II

Animals and phenotype

Data in the experiment II were collected as part of Work Package 4 from the Genotype plus Environment (**GplusE**) FP7-Project (http://www.gpluse.eu). This experiment was done on 4,267 second-parity Holstein cows with calving year 2015 to 2018, distributed over 50 herds in Belgium and the Netherlands. De Koster *et al*. (2019) compared different models to predict the metabolic clusters of early lactation dairy cows using milk metabolites or Fourier-transform mid-infrared (FT-MIR) spectra of milk and reported that the use of FT-MIR spectra had a high accuracy to predict metabolic clusters of individual cows. A detailed description of the procedure is explained in De Koster *et al*. (2019). In short, milk samples were collected starting from the first week in milk until 50 days in milk (DIM) and FT-MIR spectra of samples were determined. Then, FT-MIR spectra were used to divide animals into

either BALANCED (cow with a favourable metabolic profile; n=1,201, 28%) or the OTHER group (n=3,066, 72%).

Genotypic data

Individuals (n=31,895) were genotyped using the BovineLD (n=20,462), Bovine SNP50K (n=10,638) or BovineHD SNP panel (795 animals). Genotypes of animals were imputed to HD with a reference population of 795 HD individuals (46 males and 749 females) using FImpute V2.2 software (Sargolzaei *et al*., 2014). In total 12,367 out of 31,895 genotyped individuals, had either phenotypic data or were in the pedigree file which were used in the subsequent association analysis (the number of animals with records was 4,267, the number of animals with records and with genotypic data was 3,725, the number of animals with records and no genotypes was 542, and the number of animals with genotypes and no records was 8,642). Quality control was performed on the imputed data and SNPs with minor allele frequency less than 5% were excluded. After genomic data quality control, 566,345 out of 730,539 SNPs were available for the association analysis.

Association between the predicted metabolic clusters and production performance

The association of 305-d milk yield and the predicted metabolic clusters was determined using a linear mixed model through the inclusion of herd-year-season of calving (**HYS**), metabolic clusters (BALANCED and OTHER), covariate effects of age at first calving (**FCA**) in both linear and quadratic forms, and random effect of dam's sire. The association of production traits including daily milk yield, daily fat- and protein- corrected milk (**FPCM**)

yield, milk fat percentage, milk protein percentage, daily fat yield, daily protein yield, daily fat + protein yield, fat to protein ratio, and milk somatic cell score (**SCS**) during the entire lactation with the predicted metabolic clusters was investigated using linear mixed models. Linear mixed models were constructed with week postcalving as a repeated observation within the random factor cow. Interaction effects were removed from the model if nonsignificant ($P > 0.05$). Multiple testing correction was done using Tukey's post hoc test and significance was declared at $P < 0.05$ (R Core Team, 2017).

Variance component estimation

Pedigree information was collected for all phenotyped animals and contained a total of 43,181 individuals (12,367 and 4,267 out of 43,181 animals had genotype and phenotype data, respectively). The variance components were estimated by Bayesian inference, considering a linear single trait animal model. The linear model included HYS as systematic effect, covariate effects of FCA in both linear and quadratic forms, DIM and milk yield at sampling as well as animal direct genetic effect and residual effect as random effects. The complete model can be represented as follows:

$$y_{ij} = \mu + HYS_i + b_1(FCA_j) + b_2(FCA_j)^2 + b_3(DIM_j) + b_4(milk_j) + a_j + e_{ij}$$

where y_{ij} represents the response variable of animal j (BALANCED vs. OTHER metabolic cluster), μ is the overall mean, HYS_i is the fixed effect of i^{th} herd-year-season of calving, b_1 and b_2 are the linear and quadratic regression coefficients of the dependent variable on the age at first calving, FCA_j is the age at first calving of j^{th} cows, b_3 and b_4 are the linear regression coefficients of the dependent variable on the DIM and milk yield at sampling,

244 DIM_j and $milk_j$ are, respectively, DIM and milk yield at sampling on j^{th} cow, a_j is the additive
245 genetic effect, and e_{ij} is the random residual error. The additive genetic and residual
246 variances were obtained as follows:

247 $$\text{var}\begin{bmatrix}\mathbf{a}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{H}\sigma_a^2 & 0\\ 0 & \mathbf{I}\sigma_e^2\end{bmatrix}$$

248 where σ_a^2, and σ_e^2 are, respectively, total additive genetic and residual variances, **a** is the
249 vector of direct additive genetic effects, **e** is a vector of residual effects, and **H** is a matrix
250 that combines pedigree and genomic relationships, and its inverse consists on the integration
251 of additive and genomic relationship matrices, **A** and **G**, respectively (Aguilar et al., 2010):

252 $$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix}0 & 0\\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}\end{bmatrix}$$

253 where **A** is the numerator relationship matrix based on pedigree for all animals; $\mathbf{A}_{22}$ is the
254 numerator relationship matrix for genotyped animals; and **G** is the genomic relationship
255 matrix which was obtained using following function described by VanRaden (2008).

256 $$\mathbf{G} = \frac{\mathbf{ZDZ}'}{\sum_{i=1}^{M} 2p_i(1-p_i)}$$

257 where **Z** is a matrix of gene content adjusted for allele frequencies (0, 1 or 2
258 for *aa*, *Aa* and *AA*, respectively); **D** is a diagonal matrix of weights for SNP variances
259 (initially **D** = **I**); M is the number of SNPs, and p_i is the minor allele frequency of i^{th} SNP.
260 The **H** matrix was built scaling **G** based on $\mathbf{A_{22}}$ considering that the average of diagonal of
261 **G** is equal to average of the diagonal of $\mathbf{A_{22}}$ and, the average of off-diagonal **G** is equal to
262 average off-diagonal $\mathbf{A_{22}}$. The analysis, consisted of a single chain of 350,000 cycles with a
263 burn-in of 100,000 cycles, taking a sample every 50 iterations, was performed using

264 THRGIBBS1F90 (Misztal *et al*., 2002). Chain convergence was assessed by visual
265 inspection. The posterior estimates were obtained retrospectively using the POSTGIBBSF90
266 program (Misztal *et al*., 2002).
267
268 **Single step genome-wide association study (ssGWAS)**
269 The analyses were performed using the weighted single-step GWAS methodology *(Wang et*
270 *al*., 2012) considering the same linear animal model used to estimate the (co)variance
271 components before mentioned. The animal effects were decomposed into those for
272 genotyped ($\mathbf{a_g}$) and ungenotyped animals ($\mathbf{a_n}$). The animal effects of genotyped animals are
273 a function of SNP effects, $a_g = Zu$, where **Z** is a matrix relating genotypes of each locus and
274 **u** is a vector of SNP marker effect. The variance of animal effects was assumed as:

275 $$\mathbf{Var(a_g) = Var\ (Zu) = ZDZ'\sigma_u^2 = G^*\sigma_a^2}$$

276 Where **D** is a diagonal matrix of weights for variances of markers (**D** = **I** for GBLUP), σ_u^2 is
277 the genetic additive variance captured by each SNP marker when the weighted relationship
278 matrix (**G***) was built with no weight.
279 The SNP effects were obtained using following equation:

280 $$\mathbf{\hat{u} = \lambda DZ'\ G^{*-1}\hat{a}_g = DZ'[ZDZ']^{-1}\hat{a}_g}$$

281 where λ was defined by VanRaden (2008) as a normalizing constant, as described below

282 $$\boldsymbol{\lambda} = \frac{\boldsymbol{\sigma_u^2}}{\boldsymbol{\sigma_a^2}} = \frac{\mathbf{1}}{\sum_{\mathbf{i=1}}^{\mathbf{M}} \mathbf{2p_i(1-p_i)}}$$

283 M is the number of SNP and p_i is the frequency of the second allele in the i^{th} SNP. The
284 following iterative process described by Wang *et al*. (2012) was used to estimate the SNP

effects. 1. **D** = **I** in the first step; 2. To calculate the **G** matrix; 3. To calculate GEBVs for entire data set using ssGBLUP; 4. To convert GEBVs to SNP effects (û): $\hat{\mathbf{u}} = \boldsymbol{\lambda}\mathbf{DZ}'\,\mathbf{G}^{*-1}\hat{\mathbf{a}}_{\mathbf{g}}$; 5. To calculate the variance of each SNP: $\mathbf{d_i} = \hat{\mathbf{u}}_{\mathbf{i}}^{\mathbf{2}}\mathbf{2p_i}(\mathbf{1} - \mathbf{p_i})$, where i is the ith SNP; 6. To normalize SNP weights to remain the total genetic variance constant; exit or loop to step 2. The effects of markers were obtained by 2 iterations from step 2 to 6 as proposed by Wang *et al*. (2012). Percentage of genetic variance explained by ith region has been calculated as below:

$$\frac{\mathrm{Var}(a_i)}{\sigma_a^2} \times 100\% = \frac{\mathrm{Var}(\sum_{j=1}^{n} Z_j \hat{u}_j)}{\sigma_a^2} \times 100$$

Where a_i is genetic value of the ith region that consists of n consecutive SNP (n = 1, 5, 10, 15, 20, and 25), σ_a^2 is the total genetic variance, **Z**j is vector of SNP content of the jth SNP for all individuals, and ù$_j$ is marker effect of the jth SNP within the ith region. The results were presented by the proportion of variance explained by each window of n consecutive SNP.

RESULTS AND DISCUSSION

EXPERIMENT I

The cows (n=105) were divided either in BALANCED (n=42, 40%) and OTHER (n=63, 60%) metabolic clusters using k-means clustering based on the blood glucose, IGF-1, BHB and FFA concentration at 14 and 35 DIM. The cows were genotyped for a total of 777,962 SNPs, of which 576,092SNPs passed all quality control measures and were used for the GWAS analysis.

306 Manhattan and Q-Q plots of SNPs associated with the predicted metabolic clusters along
307 with those for the blood metabolites (concentrations of BHB, glucose, IGF-1, and FFA) used
308 to create the clusters, are presented in Figure 1. Although there was no SNP associated with
309 the predicted metabolic clusters at the 5% genome-wide significance threshold, there were
310 eight SNPs (BTA11 (n=2), BTA23 (n=2), BTAX (n=4)) associated at the suggestive
311 genome-wide significance threshold. The identified SNPs in BTA11 and BTAX were also
312 associated with IGF-1 at the suggestive genome-wide significance threshold (Figure 1). The
313 identified SNPs along with their position, and 100 kb flanking genes are presented in Table
314 1 (Assembly UMD3. 1, annotation release 103). The SNPs identified on BTA11 were
315 mapped inside QTLs for Immunoglobulin G (IgG) level (Maltecca et al., 2009), body weight
316 (McClure *et al*., 2010), and milk alpha-lactalbumin percentage (Zhou *et al*., 2019). The
317 immune response profile in transition dairy cows has been addressed in a number of studies
318 (Herr *et al*., 2011, Sordillo, 2016). Herr *et al*. (2011) reported a significant decrease of the
319 serum IgG level beginning at the 8th week before parturition and recovering by the 4th week
320 post parturition. The SNP BovineHD1100022178 on BTA11 was mapped inside the
321 *MIR2285AF-1* gene which is a microRNA (miRNA). It has been reported that miRNAs have
322 important effects on mammary gland development (Piao and Ma, 2012). The SNPs detected
323 on BTA23, mapped inside genes including *LOC101906324*, *NEDD9*, *SMIM13*, and
324 *LOC104969831*, overlap with QTLs for milk fatty acid content (Bouwman *et al*., 2012),
325 body weight (Michenet *et al*., 2016), and dry matter intake (Sherman *et al*., 2009). Linkage
326 disequilibrium (**LD**) analysis was performed between 26 SNPs (on BTAX position 17.2-
327 17.3 Mb) including four SNPs associated with the predicted metabolic clusters. The

328 estimated pair-wise r^2 values showed that the SNPs identified as associated SNPs for the
329 predicted metabolic clusters are highly correlated (Figure 2). In conclusion, the aim of this
330 experiment was to use GWAS to identify genetic markers linked with metabolic clusters in
331 early-lactation Holstein cows. However, it seems that a bigger sample size is needed for
332 detecting significant SNPs in a GWAS. Although, k-means clustering based on blood
333 metabolites can be considered as an interesting method to predict metabolic clusters of
334 transition dairy cows, the application of this method on a large sample size in order to
335 perform an accurate GWAS analysis is very challenging and expensive. De Koster *et al.*
336 (2019) suggested a method of clustering early-lactation cows based on FT-MIR spectra of
337 milk to predict metabolic clusters of animals, which was used in experiment II
338
339 **EXPERIMENT II**
340 In experiment II, milk samples were used to determine FT-MIR spectra and subsequently
341 predict the metabolic clusters for each animal. The cows (n=4,267) were grouped in either
342 BALANCED (n = 1,201, 28%) or OTHER (n = 3,066, 72%) metabolic cluster. De Koster *et*
343 *al.* (2019) reported that the use of FT-MIR spectra had a high accuracy (80%) to predict
344 metabolic clusters of individual cows. Therefore, BALANCED cows were described having
345 higher concentrations for blood glucose and IGF-I and lower concentration for blood free
346 fatty acids and BHB compared with OTHER cows. There was a significant association
347 between the 305-d milk yield and the metabolic clusters ($P< 0.05$). The average 305-d milk
348 yield (±SE) for cows belonging to the BALANCED metabolic cluster versus those belonging
349 to the OTHER metabolic cluster was 9,622 (±51.8) and 9,840 (36.2) kg, respectively. Daily

350 milk and protein yields were lower in cows from the BALANCED metabolic cluster
351 compared with those belonging to the OTHER metabolic cluster. Daily fat yield, daily fat
352 and protein yield, fat to protein ratio, milk fat percentage, and milk protein percentage were
353 higher in BALANCED cows cluster compared with those belonging to the OTHER
354 metabolic cluster, however; there was no difference for milk SCS in BALANCED versus
355 OTHER cows (Figure 3; Table 2). Although there was no difference for daily FPCM during
356 the entire lactation, FPCM was higher during early lactation in the BALANCED cows
357 compared with that for those in the OTHER metabolic cluster (Table 2; Figure 3). This is in
358 contrast to the original clustering study (de Koster et al., 2019) which reported no difference
359 in FPCM in BALANCED cows (n = 43; mean = 38.42 kg) when compared with those
360 belonging to the OTHER cluster (n = 64; mean = 37.76 kg). van Hoeij et al. (2019), reported
361 that better metabolic status, as indicated by plasma metabolites, in dairy cows in early
362 lactation is associated with a greater DMI and EB. Beerda et al. (2004) reported no
363 significant difference for the adaptive capacity in high-producing cows compared with those
364 with low-producing.
365 The posterior mean (SD) of heritability estimates for predicted metabolic clusters was 0.17
366 (0.03), indicating that the genetic variation for the predicted metabolic clusters is large
367 enough to change it through genetic selection. The heritabilities of DMI and energy balance
368 (EB) over the entire period of lactation in Nordic Red dairy cattle have been reported to be
369 0.23 and 0.10, respectively (Liinamo *et al*., 2012). Spurlock *et al*. (2012) reported heritability
370 estimates ranging from 0.27 to 0.63 and 0.12 to 0.49, respectively, for DMI and EB in
371 Holstein cows.

In the present study, windows-based GWAS was used to identify genetic marker(s) linked with metabolic clusters predicted using FT-MIR of milk in early-lactating Holstein dairy cows. The absence of a universal approach for hypothesis testing is an important challenge of window-based GWAS (Aguilar *et al*., 2019), even though it is quite a common procedure in genetic studies. In window-based GWAS, different window types (distinct or sliding windows) and variable window sizes (defined as the number of SNPs, or the number of base pairs) can be used. The common form for declaring importance is to use a threshold on the additive genetic variance explained by individual windows (Aguilar *et al*., 2019). However, it is unclear what window size is optimal and there is no standard to define the threshold on explained genetic variance (Aguilar *et al*., 2019). Therefore, determining the proper window size is usually subjective and researchers often have not justified their choices or sometimes have acknowledged that their choices are arbitrary (Myles *et al*., 2008, Beissinger *et al*., 2015). Han and Peñagaricano (2016) considered 1.5 Mb SNP windows that explained more than 0.50% of the genetic variance as threshold to declare significance. Suwannasing *et al*. (2018) using Porcine SNP60k BeadChip, considered 5-adjacents SNP windows that explained more than 1% of the total genetic variance as threshold to declare significance. De Oliveira Silva *et al*. (2017) using BovineHD SNP panel, considered 50-adjacents SNPs windows (with an average of 280 kb) that explained more than 0.5% of additive genetic variance as threshold to declare significance. Fragomeni *et al*. (2014) reported that small and large window sizes are accompanied, respectively, with large noise and absence of peaks. Beissinger *et al*. (2015) reported that small sliding windows had the most favorable ratio of detection rate to false-positive rate in comparison to large window sizes. In the present study,

394 sliding windows of 1, 5, 10, 15, 20, and 25 consecutive SNPs were used to identify genetic
395 marker(s) associated with the metabolic clusters in early-lactation Holstein dairy cows and
396 to determine if the region(s) identified may change depending on the window size (Figure
397 4). The results of single SNP windows showed small individual SNP variances for the
398 predicted metabolic clusters (with mean and standard deviation of 1.3×10^{-4} and 6.6×10^{-8},
399 respectively), with just 16 individual SNPs explaining more than 0.005% of the additive
400 genetic variance (35 times more than the mean). The top 16 individual SNPs combined
401 explained more than 0.09 % of the total additive variance and were mapped on BTA27 in
402 position 36,258 to 36,295 kb (a region with length < 38,000 bp). The mean genetic variance
403 explained by an individual 5-adjacents SNP window was 8.4×10^{-4}. The first and second top
404 5-adjacents SNP windows explained more than 0.04 and 0.03% of the genetic variance and
405 were mapped on BTA27 in position 36,265 to 36,271, and 36,259 to 36,264 kb, respectively.
406 The first and second top 10-adjacents SNP windows explained more than 0.07 and 0.06% of
407 the genetic variance and were mapped on BTA27 in position 36,258 to 36,268, and BTA20
408 in position 54,493 to 54,508 kb, respectively. The first and second top 15-adjacents SNP
409 windows explained more than 0.11 and 0.09% of the genetic variance and were mapped on
410 BTA27 in position 36,258 to 36,280, and BTA20 in position 54,485 to 54,207 kb,
411 respectively. The first top 20-adjacents SNP window explained more than 0.15% of the total
412 additive genetic variance and was mapped on BTA27 in position 36,258 to 36,290 kb. The
413 second top 20-adjacents SNP window was mapped on BTA20 in position 54,470 to 54,520
414 kb and explained around 0.09% of the total additive genetic variance. The first top 25-
415 adjacents SNP window explained more than 0.15% of the additive genetic variance and was

mapped on BTA27 in position 36,258 to 36,296 kb (a region with length < 38,000 bp). All the top 16 individual SNP identified in the single SNP analysis were presented in this window. The second top 25-adjacents SNP window was mapped on BTA7 in position 64,537 to 64,568 kb and explained 0.11% of additive genetic variance. The results of different sliding window sizes showed that the region identified on BTA27 in position 36,258 to 36,295 kb overlaps among all windows sizes considered and can be suggested to be associated with the predicted metabolic clusters in early-lactation Holstein cows. The identified region overlaps with QTL regions reported for meat quality (Horodyska *et al*., 2015), milk fatty acid profiles (Bouwman *et al*., 2011) and dry matter intake in cattle (Tetens *et al*., 2014). The identified region was mapped inside *ANK1* and *miR-486* genes. The association between *ANK1* and *miR-486* genes with the predicted metabolic clusters may be partly explained by their regulating roles in hematopoiesis (Sureshchandra *et al*., 2016), erythropoiesis (Wang *et al*., 2015b), and mammary gland development (Li *et al*., 2015). *ANK1* is a large (~240 kb) gene that encodes the adapter protein ankyrin-1. The ankyrins act as adaptors among a variety of integral membrane proteins and the spectrin skeleton (Bennett and Baines, 2001), and play important roles in cell motility, proliferation and activation (Nelson and Lazarides, 1984). Ankyrin1, the prototype of the ankyrins, has been reported to be a hematopoiesis specific regulator (Sureshchandra *et al*., 2016). In human, genetic variants located within *ANK1* have been reported to be associated with glycemic traits, impaired insulin release and type 2 diabetes (Imamura *et al*., 2012, Morris *et al*., 2012). In cattle, *ANK1* gene has been reported to be associated with meat quality, milk protein and rump conformation (Kolbehdari *et al*., 2008, Horodyska *et al*., 2015, Cai *et al*., 2018). Tetens

438 *et al*. (2014) reported a QTL for DMI, as a correlated trait with metabolic adaptation in dairy
439 cows, mapped on *ANK1*. SNPs in the *ANK1* promoter have also been reported to be
440 associated with intramuscular fat in bovine and porcine tissues *(Aslan et al., 2010, Aslan et*
441 *al*., 2012). *miR-486* is a microRNA (miRNA). miRNAs are small (~22 nucleotides) non-
442 coding RNAs that regulate many fundamental biological processes primarily through
443 affecting both the stability and translation of mRNAs (Bartel, 2004). *miR-486* is located
444 within the last intron of *ANK1* that is common to all transcripts and could be co-transcribed
445 with any or all variants (Tessema *et al*., 2017). There is evidence indicating potential co-
446 transcription and co-regulation of *ANK1* and *miR-486* (Gallagher *et al*., 2000, Shaham *et al*.,
447 2015, Tessema *et al*., 2017). In humans, biological links have been reported between
448 the *miR-486* locus and metabolic disorders including type 2 diabetes and obesity (Valsesia
449 *et al*., 2019). A significant association between residual feed intake and the expression
450 level of *miR-486* has been reported in cattle (De Oliveira *et al*., 2018) and pig (Jing *et al*.,
451 2015). Mammary gland development is controlled by multiple genes including miRNAs
452 (Piao and Ma, 2012). Li *et al*. (2015) reported that *miR-486* is expressed in both bovine
453 mammary gland tissues and in mammary epithelial cells and regulates lactation. *miR-486*
454 was also identified as a downstream regulator of PTEN (phosphatase and tensin homolog)
455 that is required for the development of the cow mammary gland (Li *et al*., 2015).
456

457 **CONCLUSION**

458 The aim of the present study was to detect genomic region(s) associated with metabolic
459 clusters in early-lactation Holstein cows. Blood metabolites (experiment I) and Fourier-

460 transform mid-infrared (FT-MIR) spectra of milk (experiment II) were used to predict
461 metabolic clusters (BALANCED vs. OTHER) for, respectively, 105 and 4,267 Holstein
462 dairy cows. The results of experiment I revealed eight SNPs (on BTA11, BTA23 and BTAX)
463 associated with the predicted metabolic clusters at the suggestive genome-wide significance
464 threshold. Considering the results of experiment II, it can be concluded that metabolic
465 clusters predicted based on FT-MIR spectra in milk samples collected during early lactation
466 is a highly polygenic trait regulated by many small-sized effects. The region of 36,258 to
467 36,295 kb on BTA27 was the highly associated region for the predicted metabolic clusters,
468 with the closest genes to this region (*ANK1* and*miR-486*) being related to hematopoiesis,
469 erythropoiesis, and mammary gland development and having been reported to be related to
470 DMI and RFI in cattle and swine. The heritability (0.17) of the predicted metabolic clusters
471 indicates that its genetic variation is large enough to alter it through genetic selection. The
472 SNPs of importance detected from this study, could be used to provide weight information
473 to the SNPs for future genomic prediction for the metabolic clusters.
474

475 **Conflict of interest**

476 The authors declare no potential conflict of interest associated with this research.
477

478 **Availability of data and material**

479 Relevant information supporting the results not presented in the manuscript is given in
480 additional files (S1 to S6 excel files).
481

Additional file:

S1 File. Results of single SNP windows (singleSNP.xlsx).

S2 File. Results of 5-adjacents SNP windows (5- adjacents.xlsx).

S3 File. Results of 10-adjacents SNP windows (10- adjacents.xlsx).

S4 File. Results of 15-adjacents SNP windows (15- adjacents.xlsx).

S5 File. Results of 20-adjacents SNP windows (20- adjacents.xlsx).

S6 File. Results of 25-adjacents SNP windows (25- adjacents.xlsx).

ACKNOWLEDGMENTS

This project has received funding from the European Union's Seventh Framework Program for research, technological development and demonstration under grant agreement no. 613689. The views expressed in this publication are the sole responsibility of the authors and do not necessarily reflect the views of the European Commission.

List of authors within the GplusE consortium: Alan Fahey, Elizabeth Matthews, Andreia Santoro, Colin Byrne, Pauline Rudd, Roisin O'Flaherty, Sinead Hallinan, Claire Wathes, Zhangrui Cheng, Ali Fouladi, Geoff Pollott, Dirk Werling, Beatriz Sanz Bernardo, Alistair Wylie, Matt Bell, Mieke Vaneetvelde, Kristof Hermans, Geert Opsomer, Sander Moerman, Jenne De Koster, Hannes Bogaert, Jan Vandepitte, Leila Vandevelde, Bonny Vanranst, Johanna Hoglund, Susanne Dahl, Soren Ostergaard, Janne Rothmann, Mogens Krogh, Else Meyer, Charlotte Gaillard, Jehan Ettema, Tine Rousing, Federica Signorelli, Francesco Napolitano, Bianca Moioli, Alessandra Crisà, Luca Buttazzoni, Jennifer McClure, Daragh Matthews, Francis Kearney, Andrew Cromie, Matt McClure, Shujun Zhang, Xing Chen,

504 Huanchun Chen, Junlong Zhao, Liguo Yang, Guohua Hua, Chen Tan, Guiqiang Wang,
505 Michel Bonneau, Andrea Pompozzi, Armin Pearn, Arnold Evertson, Linda Kosten, Anders
506 Fogh, Thomas Andersen, Matthew Lucey, Chris Elsik, Gavin Conant, Jerry Taylor, Nicolas
507 Gengler, Michel Georges, Frédéric Colinet, Hedi Hammami, Catherine Bastin, Haruko
508 Takeda, Aurelie Laine, Anne-Sophie Van Laere, Martin Schulze, Sergio Palma Vera,
509 Conrad Ferris, Cinzia Marchitelli.
510
511 **References**
512 Aguilar, I., A. Legarra, F. Cardoso, Y. Masuda, D. Lourenco, and I. Misztal. 2019.
513 Frequentist p-values for large-scale-single step genome-wide association, with an
514 application to birth weight in American Angus cattle. Genet. Sel. Evol, 51(1), 28.
515 Aguilar, I., I. Misztal, D. Johnson, A. Legarra, S. Tsuruta, and T. Lawlor. 2010. Hot topic:
516 A unified approach to utilize phenotypic, full pedigree, and genomic information for
517 genetic evaluation of Holstein final score. J. Dairy Sci. 93:743-752.
518 Alawneh, J., M. Stevenson, N. Williamson, N. Lopez-Villalobos, and T. Otley. 2012. The
519 effect of liveweight change on reproductive performance in a seasonally calving,
520 pasture fed dairy herd. Livest. Sci. 145:131-139.
521 Aslan, O., R. M. Hamill, A. M. Mullen, G. C. Davey, M. Gil, C. D. Gladney, and T.
522 Sweeney. 2012. Association between promoter polymorphisms in a key cytoskeletal
523 gene (Ankyrin 1) and intramuscular fat and water-holding capacity in porcine
524 muscle. Mol. Biol. Rep. 39:3903-3914.

525 Aslan, O., T. Sweeney, A. M. Mullen, and R. M. Hamill. 2010. Regulatory polymorphisms
526 in the bovine Ankyrin 1 gene promoter are associated with tenderness and
527 intramuscular fat content. BMC Genet. 11(1):111.
528 Bartel, D. P. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell
529 116:281-297.
530 Bauman, D. E. and W. B. Currie. 1980. Partitioning of nutrients during pregnancy and
531 lactation - a review of mechanisms involving homeostasis and homeorhesis. J. Dairy
532 Sci. 63:1514-1529.
533 Beissinger, T. M., G. J. Rosa, S. M. Kaeppler, D. Gianola, and N. De Leon. 2015. Defining
534 window-boundaries for genomic analyses using smoothing spline techniques. Genet.
535 Sel. Evol. 47(1):30.
536 Bell, A. W. 1995. Regulation of organic nutrient metabolism during transition from late
537 pregnancy to early lactation. J. Anim. Sci. 73:2804-2819.
538 Bell, A. W. and D. E. Bauman. 1997. Adaptations of glucose metabolism during pregnancy
539 and lactation. J. Mammary Gland Biol. Neoplasia 2:265-278.
540 Bennett, V. and A. J. Baines. 2001. Spectrin and ankyrin-based pathways: metazoan
541 inventions for integrating cells into tissues. Physiol. Rev. 81:1353-1392.
542 Beerda, B., J. Kornalijnslijper, J. Van der Werf, E. Noordhuizen-Stassen, and H. Hopster.
543 2004. Effects of milk production capacity and metabolic status on HPA function in
544 early postpartum dairy cows. J. Dairy Sci. 87:2094-2102.
545 Bouwman, A. C., H. Bovenhuis, M. H. Visker, and J. A. van Arendonk. 2011. Genome-wide
546 association of milk fatty acids in Dutch dairy cattle. BMC genet 12(1):43.

547 Bouwman, A. C., M. H. Visker, J. A. van Arendonk, and H. Bovenhuis. 2012. Genomic
548 regions associated with bovine milk fatty acids in both summer and winter milk
549 samples. BMC genet 13(1):93.
550 Cai, Z., B. Guldbrandtsen, M. S. Lund, and G. Sahana. 2018. Dissecting closely linked
551 association signals in combination with the mammalian phenotype database can
552 identify candidate genes in dairy cattle. BMC genet 19(1):30.
553 Collard, B., P. Boettcher, J. M. Dekkers, D. Petitclerc, and L. Schaeffer. 2000. Relationships
554 between energy balance and health traits of dairy cattle in early lactation. J. Dairy
555 Sci. 83:2683-2690.
556 Contreras, G. A., C. Strieder-Barboza, and J. De Koster. 2018. Symposium review:
557 Modulating adipose tissue lipolysis and remodeling to improve immune function
558 during the transition period and early lactation of dairy cows. J. Dairy Sci. 101:2737-
559 2752.]
560 De Koster, J., M. Salavati, C. Grelet, M. A. Crowe, E. Matthews, R. O'Flaherty, G. Opsomer,
561 L. Foldager, and M. Hostens. 2019. Prediction of metabolic clusters in early-lactation
562 dairy cows using models based on milk biomarkers. J. Dairy Sci. 102:2631-2644.
563 de Oliveira Silva, R. M., N. B. Stafuzza, B. de Oliveira Fragomeni, G. M. F. de Camargo,
564 T. M. Ceacero, J. N. d. S. G. Cyrillo, F. Baldi, A. A. Boligon, M. E. Z. Mercadante,
565 and D. L. Lourenco. 2017. Genome-wide association study for carcass traits in an
566 experimental Nelore cattle population. **PLoS ONE** 12(1):e0169860.
567 De Oliveira, P. S., L. L. Coutinho, P. C. Tizioto, A. S. Cesar, G. B. de Oliveira, W. J. d. S.
568 Diniz, A. O. De Lima, J. M. Reecy, G. B. Mourão, and A. Zerlotini. 2018. An

integrative transcriptome analysis indicates regulatory mRNA-miRNA networks for residual feed intake in Nelore cattle. Sci. Rep. 8(1):17072.

Drackley, J. K. 1999. Biology of dairy cows during the transition period: The final frontier? J. Dairy Sci. 82:2259-2273.

Drackley, J. K., H. M. Dann, G. N. Douglas, N. A. J. Guretzky, N. B. Litherland, J. P. Underwood, and J. J. Loor. 2005. Physiological and pathological adaptations in dairy cows that may increase susceptibility to periparturient diseases and disorders. Ital. J. Anim. Sci. 4:323-344.

Esposito, G., P. C. Irons, E. C. Webb, and A. Chapwanya. 2014. Interactions between negative energy balance, metabolic diseases, uterine health and immune response in transition dairy cows. Anim. Reprod. Sci. 144:60-71.

Foldager, L., M. Hostens, M. Salavati, C. Grelet, M. Sorensen, D. Wathes, C. Ferris, C. Marchitelli, F. Becker, T. Larsen, E. Matthews, R. O'Flaherty, G. Opsomer, N. Gengler, M. Crowe, GplusE Consortium, and K. Ingvartsen. 2019. Predicting risk of physiological imbalance in Holstein dairy cows from milk biomarkers. Animal.

Fragomeni, B. d. O., I. Misztal, D. L. Lourenco, I. Aguilar, R. Okimoto, and W. M. Muir. 2014. Changes in variance explained by top SNP windows over generations for three traits in broiler chicken. Front Genet. 5:332.

Gallagher, P. G., M. Romana, T. T. William, S. E. Lux, and B. G. Forget. 2000. The human ankyrin-1 gene is selectively transcribed in erythroid cell lines despite the presence of a housekeeping-like promoter. Blood 96:1136-1143.

590 Goff, J. and R. Horst. 1997. Physiological changes at parturition and their relationship to
591 metabolic disorders. J. Dairy Sci. 80:1260-1268.
592 Han, Y. and F. Peñagaricano. 2016. Unravelling the genomic architecture of bull fertility in
593 Holstein cattle. BMC genet 17(1):143.
594 Herr, M., H. Bostedt, and K. Failing. 2011. IgG and IgM levels in dairy cows during the
595 periparturient period. Theriogenology 75:377-385.
596 Horodyska, J., T. Sweeney, M. Ryan, and R. Hamill. 2015. Novel SNPs in the Ankyrin 1
597 gene and their association with beef quality traits. Meat Sci. 108:88-96.
598 Imamura, M., S. Maeda, T. Yamauchi, K. Hara, K. Yasuda, T. Morizono, A. Takahashi, M.
599 Horikoshi, M. Nakamura, and H. Fujita. 2012. A single-nucleotide polymorphism in
600 ANK1 is associated with susceptibility to type 2 diabetes in Japanese populations.
601 Hum. Mol. Genet. 21:3042-3049.
602 Ingvartsen, K. L. 2006. Feeding- and management-related diseases in the transition cow -
603 Physiological adaptations around calving and strategies to reduce feeding-related
604 diseases. Anim. Feed Sci. Tech. 126:175-213.
605 Ingvartsen, K. L. and J. B. Andersen. 2000. Integration of metabolism and intake regulation:
606 A review focusing on periparturient animals. J. Dairy Sci. 83:1573-1597.
607 Ingvartsen, K. L., R. J. Dewhurst, and N. C. Friggens. 2003. On the relationship between
608 lactational performance and health: is it yield or metabolic imbalance that cause
609 production diseases in dairy cattle? A position paper. Livest. Prod. Sci. 83:277-308.

610 Jing, L., Y. Hou, H. Wu, Y. Miao, X. Li, J. Cao, J. M. Brameld, T. Parr, and S. Zhao. 2015.
611 Transcriptome analysis of mRNA and miRNA in skeletal muscle indicates an
612 important network for differential Residual Feed Intake in pigs. Sci. Rep. 5:11953.
613 Kessel, S., M. Stroehl, H. H. D. Meyer, S. Hiss, H. Sauerwein, F. J. Schwarz, and R. M.
614 Bruckmaier. 2008. Individual variability in physiological adaptation to metabolic
615 stress during early lactation in dairy cows kept under equal conditions. J. Anim. Sci.
616 86:2903-2912.
617 Kolbehdari, D., Z. Wang, J. Grant, B. Murdoch, A. Prasad, Z. Xiu, E. Marques, P. Stothard,
618 and S. Moore. 2008. A whole-genome scan to map quantitative trait loci for
619 conformation and functional traits in Canadian Holstein bulls. J. Dairy Sci. 91:2844-
620 2856.
621 Lander, E. and L. Kruglyak. 1995. Genetic dissection of complex traits: guidelines for
622 interpreting and reporting linkage results. Nat. Genet. 11(3):241.
623 LeBlanc, S. 2010. Monitoring metabolic health of dairy cattle in the transition period. J
624 Reprod Develop. 56(S):S29-S35.
625 Li, D., X. Xie, J. Wang, Y. Bian, Q. Li, X. Gao, and C. Wang. 2015. MiR-486 regulates
626 lactation and targets the PTEN gene in cow mammary glands. PLoS ONE
627 10(3):e0118284.
628 Liinamo, A.-E., P. Mäntysaari, and E. Mäntysaari. 2012. Genetic parameters for feed intake,
629 production, and extent of negative energy balance in Nordic Red dairy cattle. J. Dairy
630 Sci. 95:6788-6794.

631 Maltecca, C., K. Weigel, H. Khatib, M. Cowan, and A. Bagnato. 2009. Whole-genome scan
632 for quantitative trait loci associated with birth weight, gestation length and passive
633 immune transfer in a Holstein× Jersey crossbred population. Anim Genet. 40:27-34.
634 McClure, M., N. Morsci, R. Schnabel, J. Kim, P. Yao, M. Rolf, S. McKay, S. Gregg, R.
635 Chapple, and S. Northcutt. 2010. A genome scan for quantitative trait loci
636 influencing carcass, post-natal growth and reproductive traits in commercial Angus
637 cattle. Anim Genet. 41:597-607.
638 Michenet, A., M. Barbat, R. Saintilan, E. Venot, and F. Phocas. 2016. Detection of
639 quantitative trait loci for maternal traits using high-density genotypes of Blonde
640 d'Aquitaine beef cattle. BMC genet 17(1):88.
641 Misztal, I., S. Tsuruta, T. Strabel, B. Auvray, T. Druet, and D. Lee. 2002. BLUPF90 and
642 related programs (BGF90). Pages 743-744 in Proc. Proceedings of the 7th world
643 congress on genetics applied to livestock production.
644 Morris, A. P., B. F. Voight, T. M. Teslovich, T. Ferreira, A. V. Segre, V. Steinthorsdottir,
645 R. J. Strawbridge, H. Khan, H. Grallert, and A. Mahajan. 2012. Large-scale
646 association analysis provides insights into the genetic architecture and
647 pathophysiology of type 2 diabetes. Nat. Genet. 44(9):981.
648 Myles, S., K. Tang, M. Somel, R. E. Green, J. Kelso, and M. Stoneking. 2008. Identification
649 and analysis of genomic regions with large between-population differentiation in
650 humans. Ann. Hum. Genet. 72(1):99-110.

651 Nelson, W. J. and E. Lazarides. 1984. Goblin (ankyrin) in striated muscle: identification of
652 the potential membrane receptor for erythroid spectrin in muscle cells. Proceedings
653 of the National Academy of Sciences 81(11):3292-3296.
654 Niculescu, L. S., N. Simionescu, E. V. Fuior, C. S. Stancu, M. G. Carnuta, M. D. Dulceanu,
655 M. Raileanu, E. Dragan, and A. V. Sima. 2018. Inhibition of miR-486 and miR-92a
656 decreases liver and plasma cholesterol levels by modulating lipid-related genes in
657 hyperlipidemic hamsters. Mol. Biol. Rep. 45(4):497-509.
658 Patton, J., D. A. Kenny, J. F. Mee, F. P. O'Mara, D. C. Wathes, M. Cook, and J. J. Murphy.
659 2006. Effect of milking frequency and diet on milk production, energy balance, and
660 reproduction in dairy cows. J. Dairy Sci. 89:1478-1487.
661 Piao, H.-l. and L. Ma. 2012. Non-coding RNAs as regulators of mammary development and
662 breast cancer. J. Mammary Gland Biol. Neoplasia. 17:33-42.
663 Pryce, J., K. P. Gaddis, A. Koeck, C. Bastin, M. Abdelsayed, N. Gengler, F. Miglior, B.
664 Heringstad, C. Egger-Danner, and K. Stock. 2016. Invited review: Opportunities for
665 genetic improvement of metabolic diseases. J. Dairy Sci. 99(9):6855-6873.
666 Puppel, K. and B. Kuczynska. 2016. Metabolic profiles of cow's blood; a review. J. Sci.
667 Food Agric. 96:4321-4328.
668 Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P.
669 Sklar, P. I. De Bakker, and M. J. Daly. 2007. PLINK: a tool set for whole-genome
670 association and population-based linkage analyses. Am. J. Hum. Genet. 81:559-575.
671 R Core Team. 2017. R: A language and environment for statistical computing. R Foundation
672 for Statistical Computing. Vienna, Austria.

673 Sargolzaei, M., J. P. Chesnais, and F. S. Schenkel. 2014. A new approach for efficient
674 genotype imputation using information from relatives. BMC Genomics 15(1):478.
675 Shaham, L., E. Vendramini, Y. Ge, Y. Goren, Y. Birger, M. R. Tijssen, M. McNulty, I.
676 Geron, O. Schwartzman, and L. Goldberg. 2015. MicroRNA-486-5p is an erythroid
677 oncomiR of the myeloid leukemias of Down syndrome. Blood 125(8):1292-1301.
678 Sherman, E., J. Nkrumah, C. Li, R. Bartusiak, B. Murdoch, and S. Moore. 2009. Fine
679 mapping quantitative trait loci for feed intake and feed efficiency in beef cattle. J.
680 Anim Sci. 87:37-45.
681 Sordillo, L. 2016. Nutritional strategies to optimize dairy cattle immunity. J. Dairy Sci.
682 99:4967-4982.
683 Spurlock, D., J. Dekkers, R. Fernando, D. Koltes, and A. Wolc. 2012. Genetic parameters
684 for energy balance, feed efficiency, and related traits in Holstein cattle. J. Dairy Sci.
685 95:5393-5402.
686 Sureshchandra, S., M. Rais, C. Stull, K. Grant, and I. Messaoudi. 2016. Transcriptome
687 profiling reveals disruption of innate immunity in chronic heavy ethanol consuming
688 female rhesus macaques. PLoS ONE 11(7):e0159295.
689 Suwannasing, R., M. Duangjinda, W. Boonkum, R. Taharnklaew, and K. Tuangsithtanon.
690 2018. The identification of novel regions for reproduction trait in Landrace and Large
691 White pigs using a single step genome-wide association study. Asian-australas. J.
692 Anim. Sci. 31:1852-1862.
693 Tessema, M., C. M. Yingling, M. A. Picchi, G. Wu, T. Ryba, Y. Lin, A. O. Bungum, E. S.
694 Edell, A. Spira, and S. A. Belinsky. 2017. ANK1 Methylation regulates expression

of MicroRNA-486-5p and discriminates lung tumors by histology and smoking status. Cancer Lett. 410:191-200.

Tetens, J., G. Thaller, and N. Krattenmacher. 2014. Genetic and genomic dissection of dry matter intake at different lactation stages in primiparous Holstein cows. J. Dairy Sci. 97:520-531.

Valsesia, A., Q.-P. Wang, N. Gheldof, J. Carayol, H. Ruffieux, T. Clark, V. Shenton, L. J. Oyston, G. Lefebvre, and S. Metairon. 2019. Genome-wide gene-based analyses of weight loss interventions identify a potential role for NKX6. 3 in metabolism. Nat. Commun. 10(1):540.

Van der Drift, S., M. Houweling, J. Schonewille, A. Tielens, and R. Jorritsma. 2012. Protein and fat mobilization and associations with serum β-hydroxybutyrate concentrations in dairy cows. J. Dairy Sci. 95:4911-4920.

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414-4423.

van Hoeij, R., A. Kok, R. Bruckmaier, M. Haskell, B. Kemp, and A. van Knegsel. 2019. Relationship between metabolic status and behavior in dairy cows in week 4 of lactation. animal 13:640-648.

Wang, H., I. Misztal, I. Aguilar, A. Legarra, and W. Muir. 2012. Genome-wide association mapping including phenotypes from relatives without genotypes. Genet. Res. 94:73-83.

715 Wang, K., D. Liu, J. Hernandez-Sanchez, J. Chen, C. Liu, Z. Wu, M. Fang, and N. Li. 2015a.
716 Genome wide association analysis reveals new production trait genes in a male
717 Duroc population. PLoS ONE 10(9):e0139207.
718 Wang, L.-S., L. Li, L. Li, S. Chu, K.-D. Shiang, M. Li, H.-Y. Sun, J. Xu, F.-J. Xiao, and G.
719 Sun. 2015b. MicroRNA-486 regulates normal erythropoiesis and enhances growth
720 and modulates drug response in CML progenitors. Blood 125:1302-1313.
721 Zhang, H., Z. Wang, S. Wang, and H. Li. 2012. Progress of genome wide association study
722 in domestic animals. J. Anim. Sci. Biotechnol. 3(1):26.
723 Zhou, C., C. Li, W. Cai, S. Liu, H. W. Yin, S. Shi, Q. Zhang, and S. Zhang. 2019. Genome-
724 wide association study for milk protein composition traits in a Chinese Holstein
725 population using a single-step approach. Front Genet. 10:72.
726 Zhou, X. and M. Stephens. 2012. Genome-wide efficient mixed-model analysis for
727 association studies. Nat. Genet .44(7):821.
728 Zwald, N., K. Weigel, Y. Chang, R. Welper, and J. Clay. 2004. Genetic selection for health
729 traits using producer-recorded data. II. Genetic correlations, disease probabilities,
730 and relationships with existing traits. J. Dairy Sci. 87(12):4295-4302.
731
732
733
734
735
736

737 **Table 1.** Single nucleotide polymorphisms (SNP) detected by genome-wide association study
738 (GWAS) for BALANCED vs. OTHER metabolic clusters[1] (Experiment I).

SNP [2]	BTA[3]	Position[4]	*P* [5]	MAF [6]	100 kb flanking genes[7]
BovineHD1100021914	11	76586236	1.8×10^{-7}	0.367 (T/C)	---
BovineHD1100022178	11	77436574	1.6×10^{-7}	0.371 (TC)	*MIR2285AF-1* (27,582), *LOC104973434* (inside), *LOC785416* (56419)
BovineHD2300013006	23	44852748	5.9×10^{-7}	0.371 (GA)	*LOC101906324* (88,276), *NEDD9* (inside)
BovineHD2300013034	23	44937247	5.9×10^{-7}	0.371 (TC)	*SMIM13* (83,227), *LOC104969831* (51,090), *NEDD9* (inside)
BovineHD3000005706	30	17232546	2.1×10^{-7}	0.252 (GA)	*MIR6526-3* (19,914), *GPC4* (37,007), *GPC3* (72,982)
BovineHD3000005716	30	17257114	3.8×10^{-7}	0.252 (AG)	*MIR6526-3* (44,482), *GPC4* (61,575), *GPC3* (48,414)
BovineHD3000005723	30	17277730	3.8×10^{-7}	0.371 (AG)	*MIR6526-3* (65,098), *GPC4* (82,191), *GPC3* (27,798)
BovineHD3000005726	30	17216072	9.6×10^{-7}	0.257 (GT)	*MIR6526-3* (3,440), *GPC4* (20,533), *GPC3* (89,456)

739 [1]BALANCED and OTHER metabolic clusters were created using k-means clustering based on log-
740 transformed and standardized concentrations of glucose, IGF-1, FFA and BHB in blood samples
741 taken from 105 multiparous Holstein dairy cows around 14 and 35 DIM.

742 [2]Single nucleotide polymorphism name.

743 [3]*Bos taurus* chromosome number

744 [4]Position of the SNP on the chromosome.

745 [5]*P* value of the mixed model for single SNP association analysis.

746 [6]Minor allele frequency and two alleles for the SNP.

747 [7]Genes in 1.0-Mb flanking regions of SNP position. Numbers in parentheses represent the physical
748 distance (in base pairs) between the SNP and the corresponding genes. Official gene symbol
749 (Assembly UMD3.1, annotation release 103).

750 **Table 2.** The association of predicted metabolic clusters[1] (BALANCED vs. OTHER) with daily milk
751 yield, fat-and protein-corrected milk (FPCM) yield, fat percentage, protein percentage, fat yield,
752 protein yield, fat + protein yield, fat to protein ratio and milk SCS during entire period of lactation.

Item	BALANCED	OTHER
Milk yield (kg/d) LSM (SE)	32.8 (1.89)[b]	33.5 (1.89)[a]
FPCM (kg/d) LSM (SE)	35.1 (1.70)[a]	34.7 (1.69)[a]
Fat percentage (%)	4.51 (0.17)[a]	4.24 (0.17)[b]
Protein percentage (%)	3.64 (0.09)[a]	3.61 (0.09)[b]
Fat yield (kg/d) LSM (SE)	1.46 (0.07)[a]	1.40(0.07)[b]
Protein yield (kg/d) LSM (SE)	1.18 (0.06)[b]	1.19 (0.06)[a]
Fat + protein yield (kg/d) LSM (SE)	2.64 (0.12)[a]	2.59 (0.12)[b]
Fat to protein ratio	1.24 (0.03)[a]	1.18 (0.03)[b]
Milk SCS	3.72(0.18)[a]	3.73 (0.18)[a]

753 [a,b] Least squares means with different superscripts differ ($P < 0.05$).
754 [1.]Milk samples collected starting from the first week in milk until 50 days in milk (DIM) were used
755 to determine Fourier-transform mid-infrared (FT-MIR) spectra and subsequently used to predict
756 metabolic clusters of animals (BLANCED vs. OTHER).
757
758
759
760
761
762
763
764
765
766
767
768
769

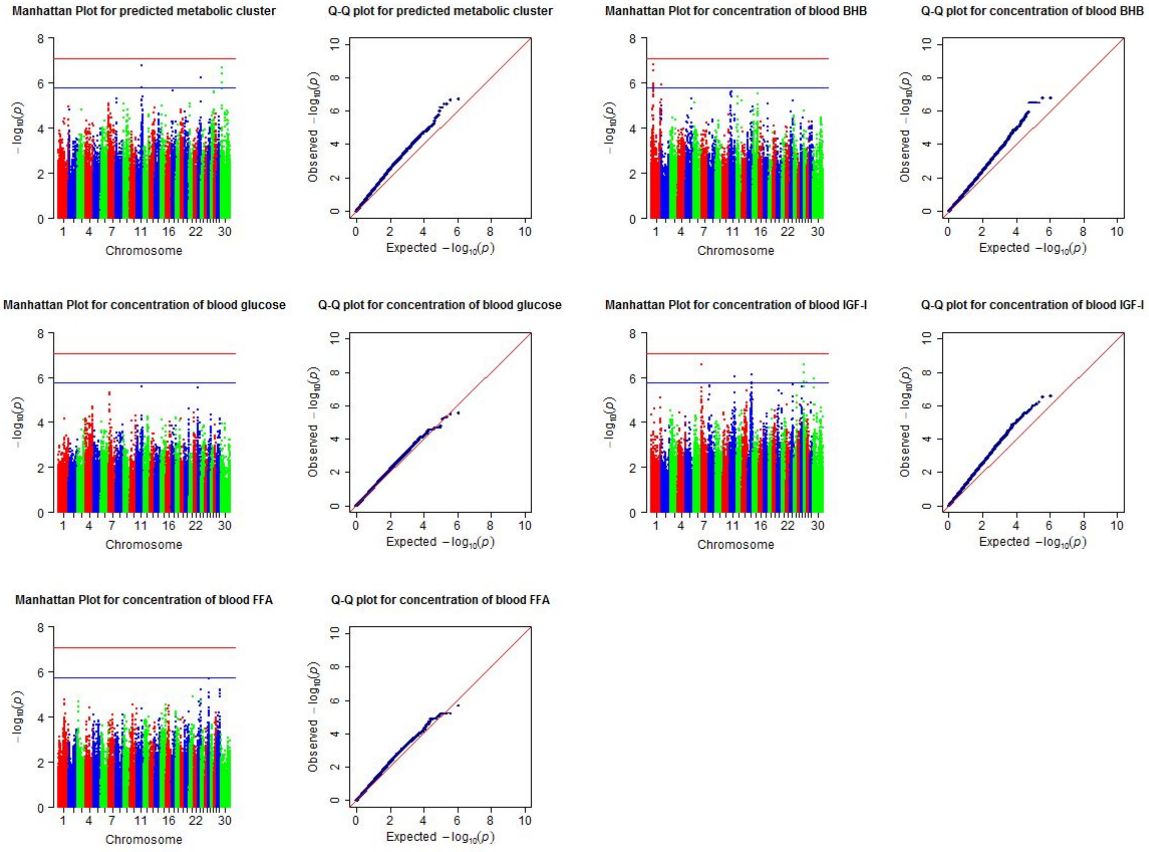

Figure 1. Manhattan and Q-Q plots of SNPs associated with the predicted metabolic clusters along with those for the metabolites (concentrations of BHB, glucose, IGF-1, and FFA and in blood samples) used to create the clusters. The horizontal red and blue lines in the Manhattan plots indicate the genome-wide significance threshold ($-\log10(8.7 \times 10^{-8})$), and suggestive significance threshold ($-\log10(1.74 \times 10^{-6})$). The metabolic clusters were created using k-means clustering based on log-transformed and standardized concentrations of glucose, IGF-1, FFA and BHB in blood samples taken from 105 multiparous Holstein dairy cows around 14 and 35 DIM (Experiment I).

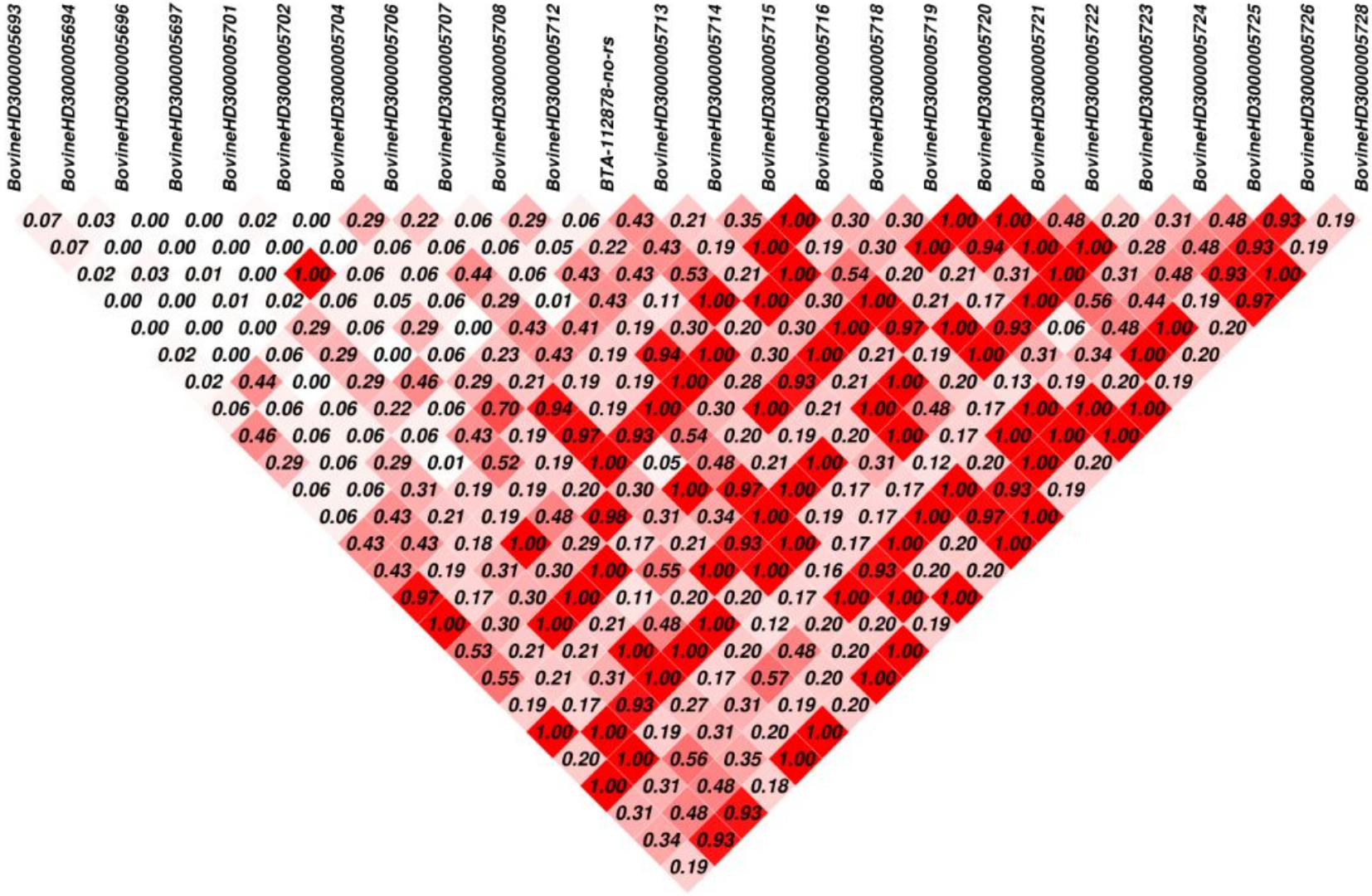

Figure 2: Linkage disequilibrium between 26 SNPs (on BTAX position 17.2-17.3 Mb) including four SNPs associated with the predicted metabolic clusters. The color scale ranges from red to white (color intensity decreases with decreasing r^2 value) (Experiment I).

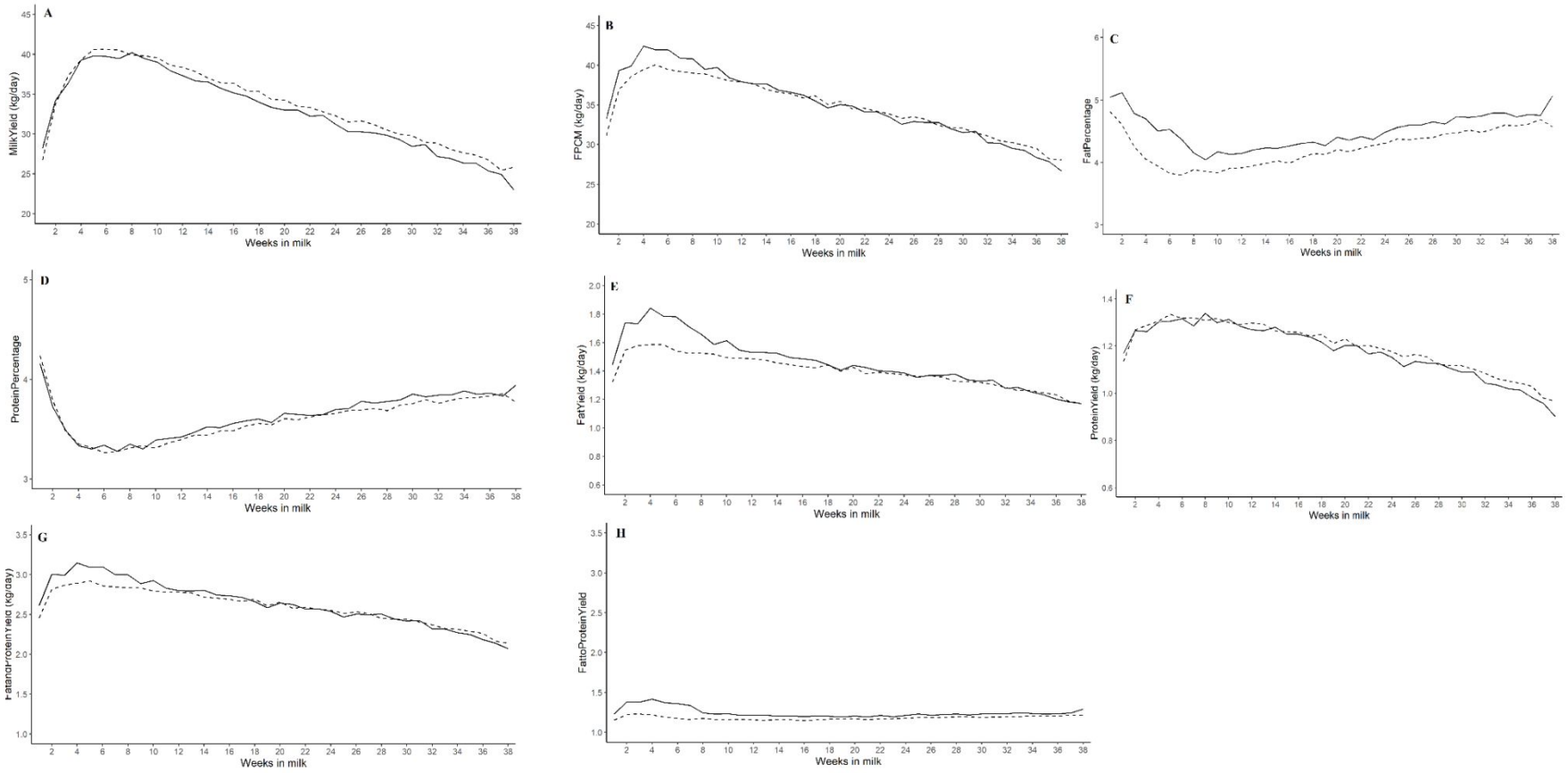

Figure 3. Daily milk yield (A), daily fat- and protein-corrected milk (FPCM) yield (B), milk fat percentage (C), milk protein percentage (D), daily fat yield (E), daily protein yield (F), daily fat+protein yield (G), and fat to protein ratio (H) in cows with a BALANCED (continuous line) and those with an OTHER (dashed line) metabolic cluster. Milk samples collected starting from the first week in milk until 50 days in milk were used to determine Fourier-transform mid-infrared (FT-MIR) spectra to classify the animals in identical metabolic clusters, BALANCED versus OTHER (Experiment II).

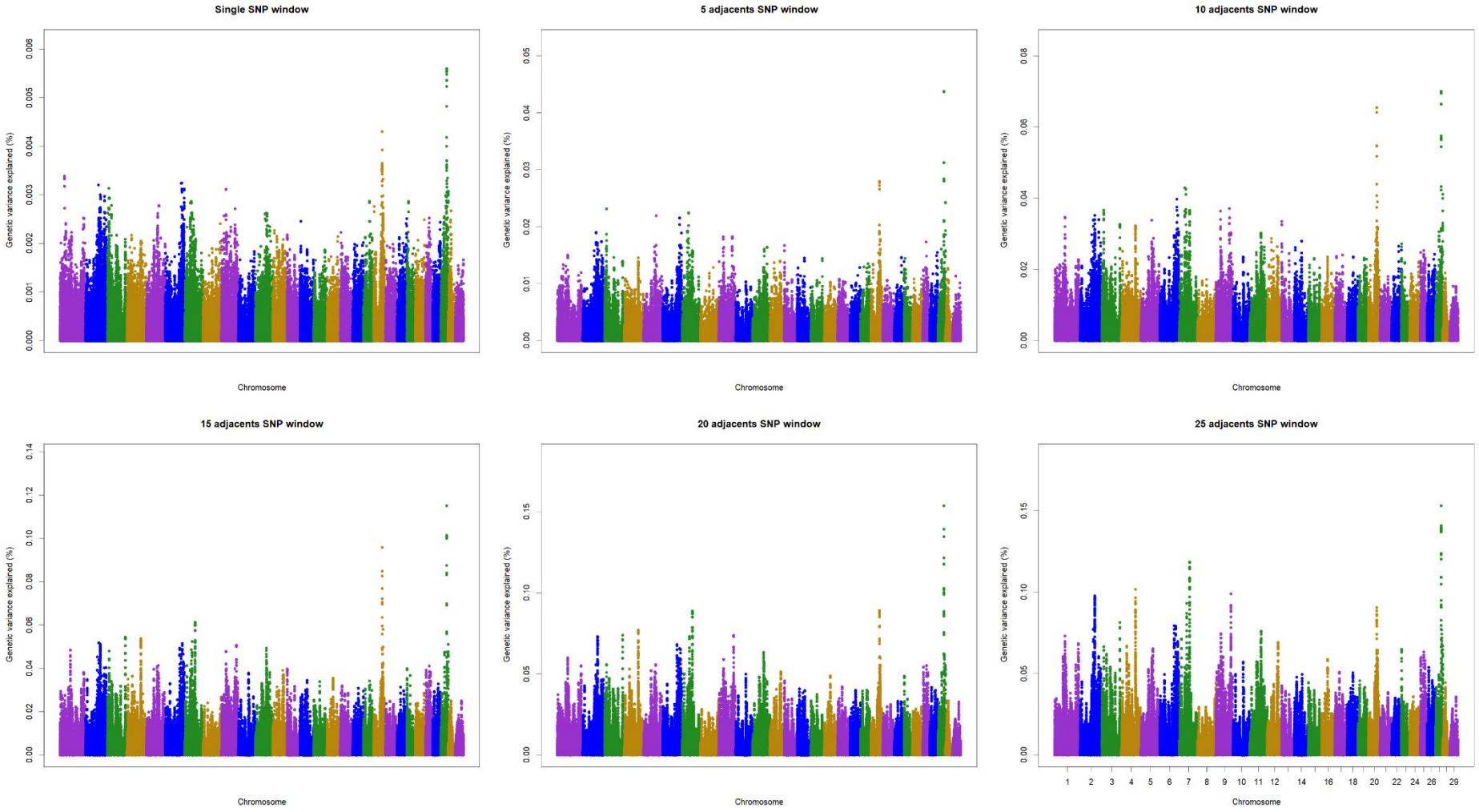

Figure 4. Additive genetic variance explained by windows of 1-, 5-, 10-, 15-, 20-, and 25-adjacents SNP across chromosomes for metabolic adaptations in early-lactation Holstein cows (Experiment II).